How does the visual brain solve its perceptual puzzles?

Daniel Kersten
Computational Vision Lab
Psychology Department, U. Minnesota
October 2008

kersten.org

Summary

- Generative knowledge used to resolve ambiguity about objects
- Some instances of perceptual “aha’s” suggest generative mechanisms
- Requires an understanding of interactions between cortical areas

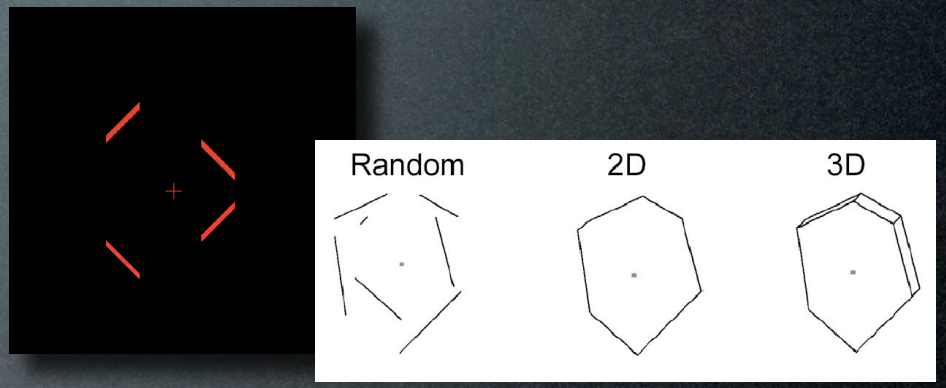
Teaser

- Oxymorons?
  - Perception as “unconscious inference”
  - “Unconscious insight?”

Outline

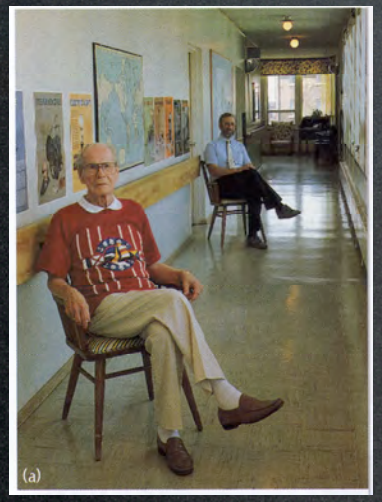
- Object perception: The problem of ambiguity
- Theory: Vision as statistical inference
- Behavior: Perceptual “puzzles”
- Mechanisms: Resolving ambiguity

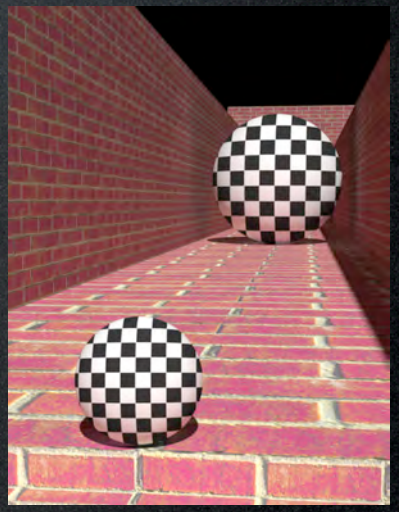
Geometric properties



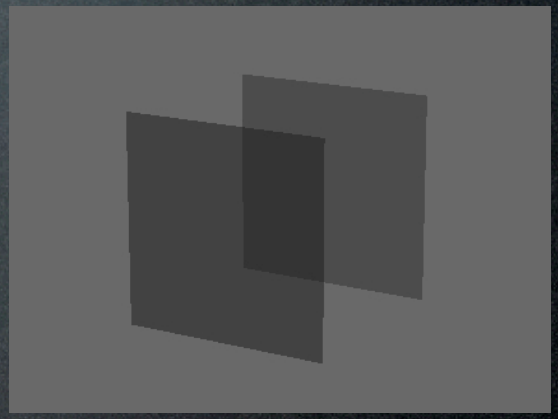
Shape

Size & depth





Material properties

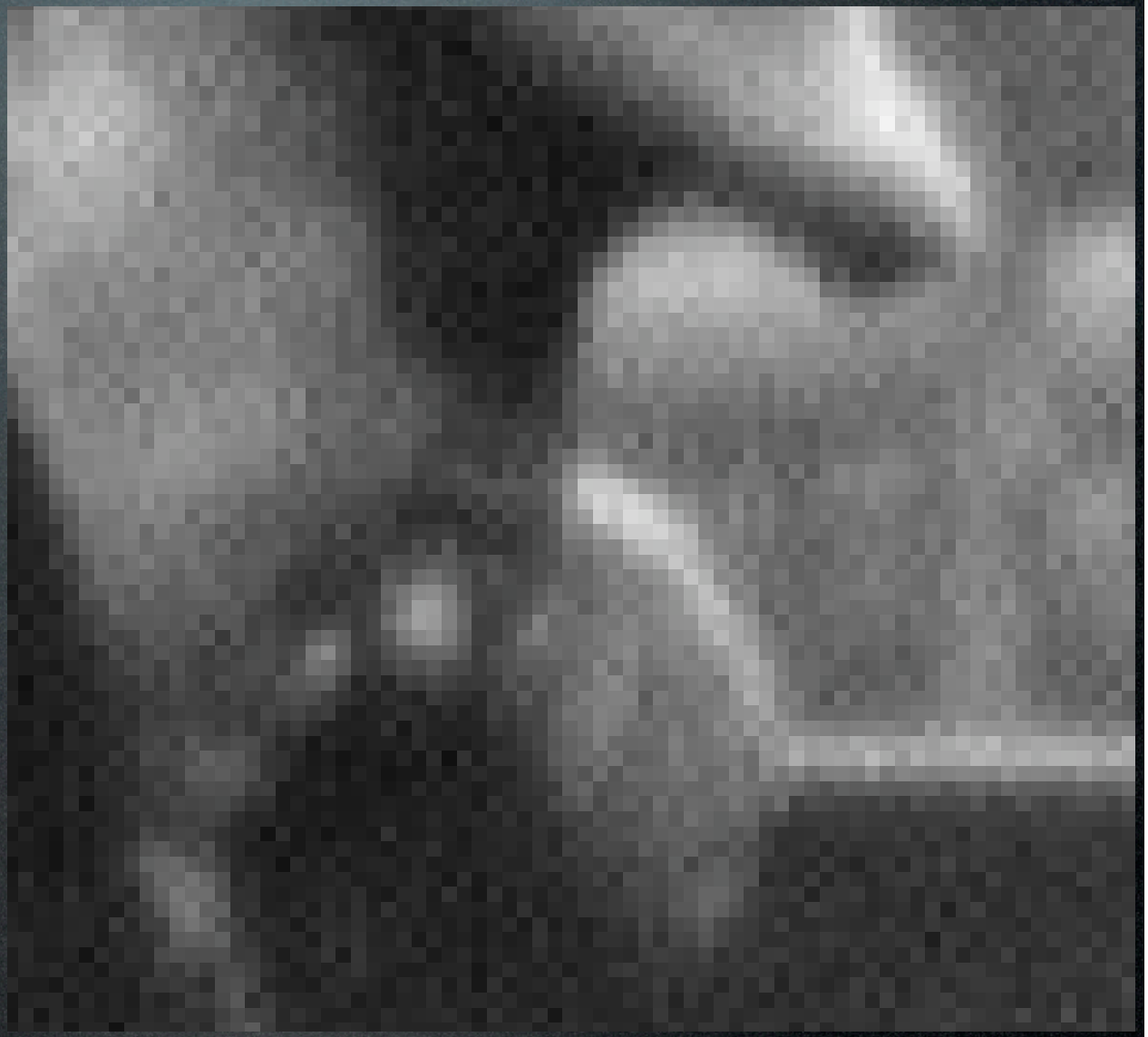


Transparency

Local ambiguity

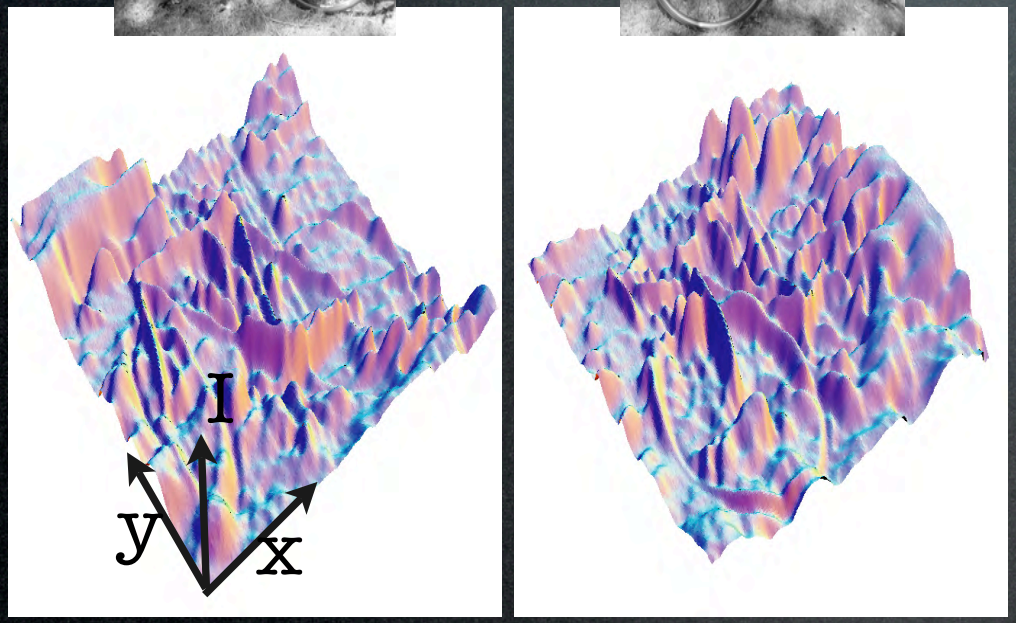
How to get reliable and useful information about objects/scene S, from complex patterns I, with locally ambiguous image data?

From: Mumford, 2002



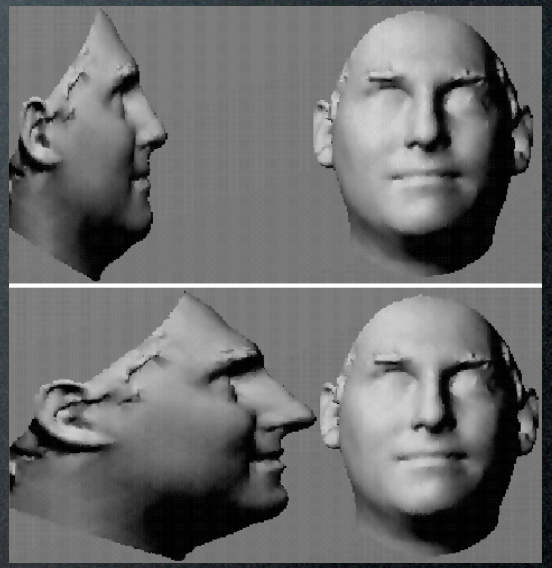
Objective ambiguity is the norm

Same object S,
different images, I_1 & I_2



Surface representation of
image intensities, I(x,y)

Different objects S_1 & S_2,
same image I



Subjective
ambiguity is rare

Outline

- Object perception: The problem of ambiguity
- Theory: Humans as Bayesian agents
- Behavior: Perceptual “puzzles”
- Mechanisms: Resolving ambiguity

Bayes tools

- Key concepts important to describing perceptual behavior
    - Represent hypotheses AND uncertainty
    - Utility
        - what to estimate with precision and what to discount
- Provide quantitative models that bridge behavior and neural networks

Humans as Bayesian agents

Actions/Decisions based on:

- models of probability on structured representations ("built-in" knowledge) of how object hypotheses S explain image patterns I: p(S, I)
- task utility, L(a(I),S)

  -> Choose actions that maximize utility, given uncertainty

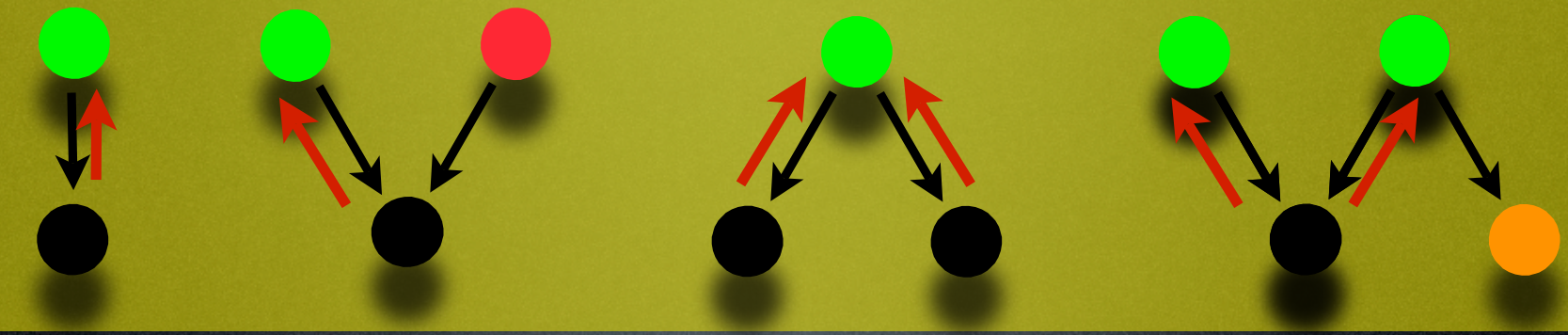
Lots of structure here!

$$a^*(I) = \arg\max_a \sum_S p(S|I) L(a(I), S)$$

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian Inference. Annual Review of Psychology, 55, 271-304.

Simple influence graphs

$p(S \mid I)$ $p(S_1 \mid I_1)$ $p(S_1 \mid I_1, I_2)$ $p(S_1, S_2 \mid I_1)$



Basic Bayes

Trade-off between prior and image "data"

Invariance/ Discounting

Cue integration

Observations are conditionally independent

Explaining away

Observation induces conditional dependence

Need to estimate accurately

Image measurement

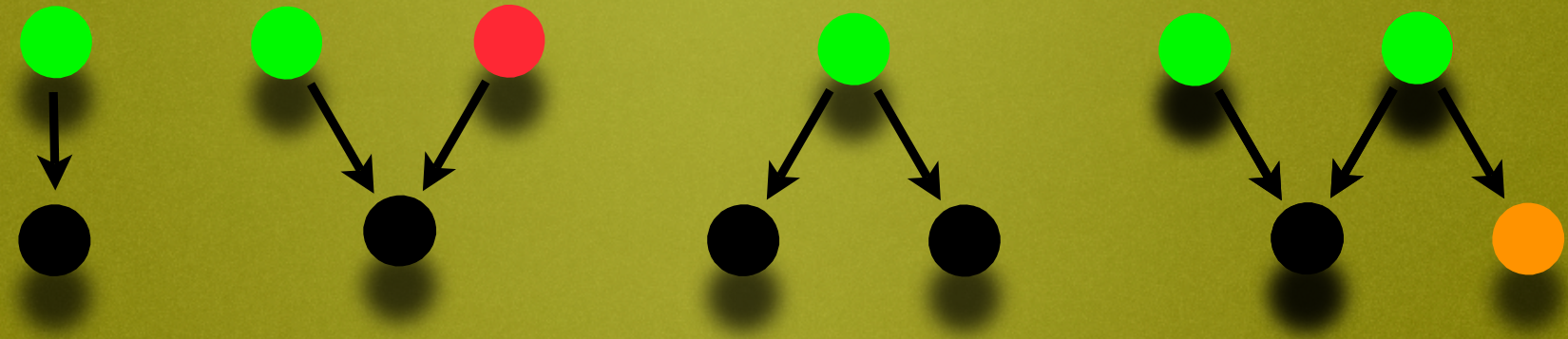
Do not need to estimate accurately

Auxiliary measurement

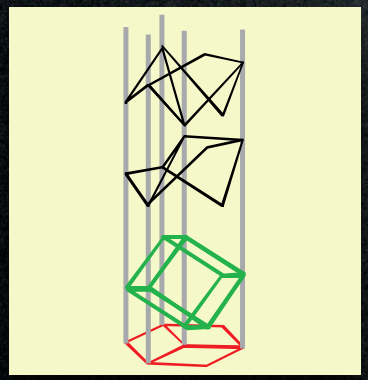
Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. Current Opinion in Neurobiology, 13(2), 1-9.

Simple influence graphs

$p(S \mid I)$ $p(S_1 \mid I_1)$ $p(S_1 \mid I_1, I_2)$ $p(S_1, S_2 \mid I_1)$



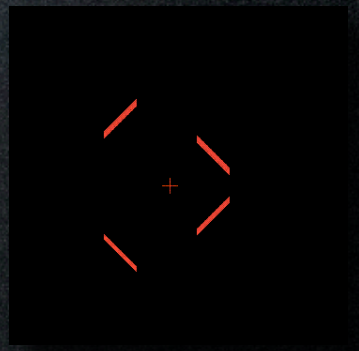
Basic Bayes



Invariance/ Discounting

Cue integration

Explaining away



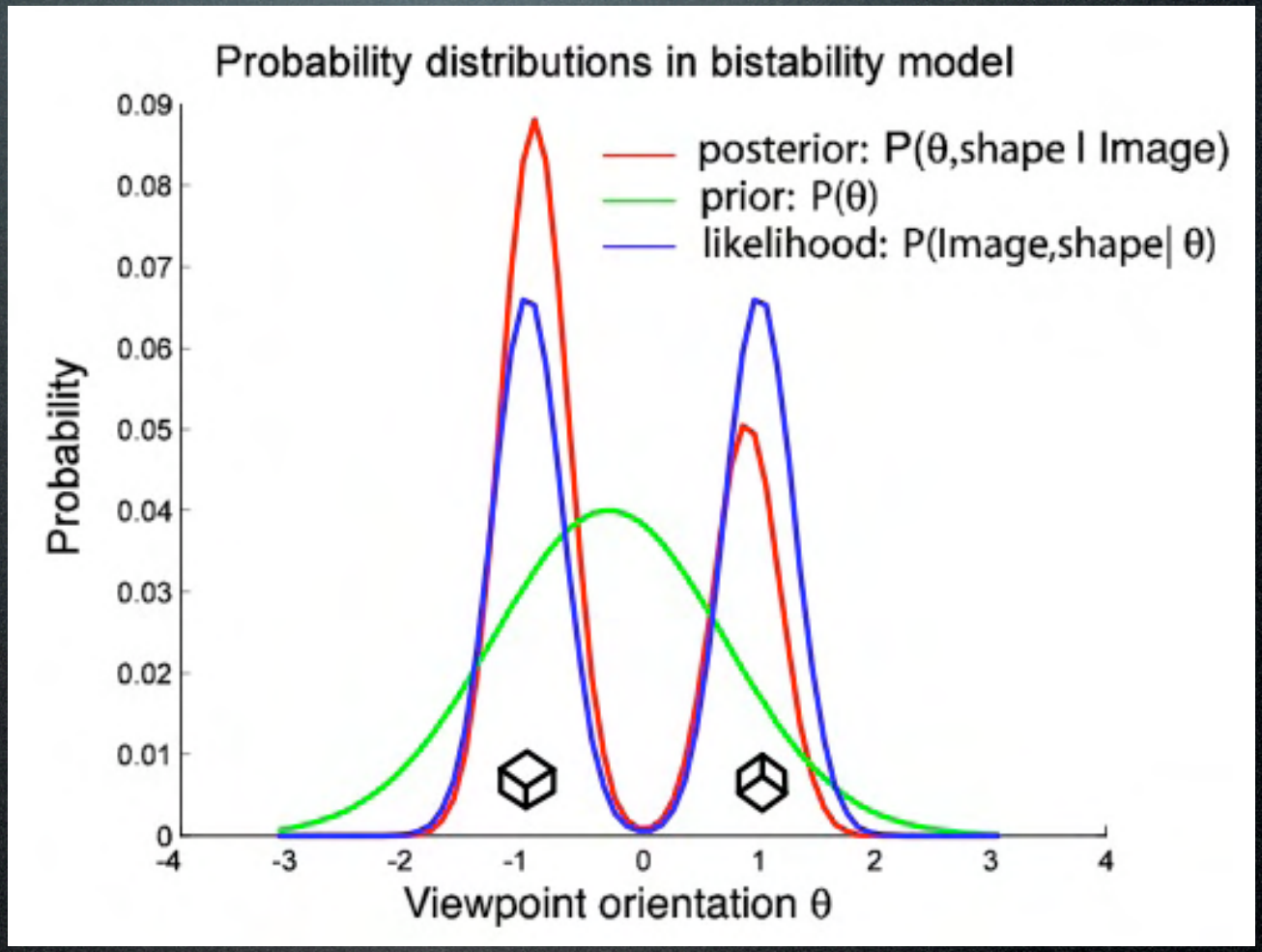
← Quantitative models →

Cf. Kersten, Mamassian & Yuille, 2004

Necker cube

p(S | I)

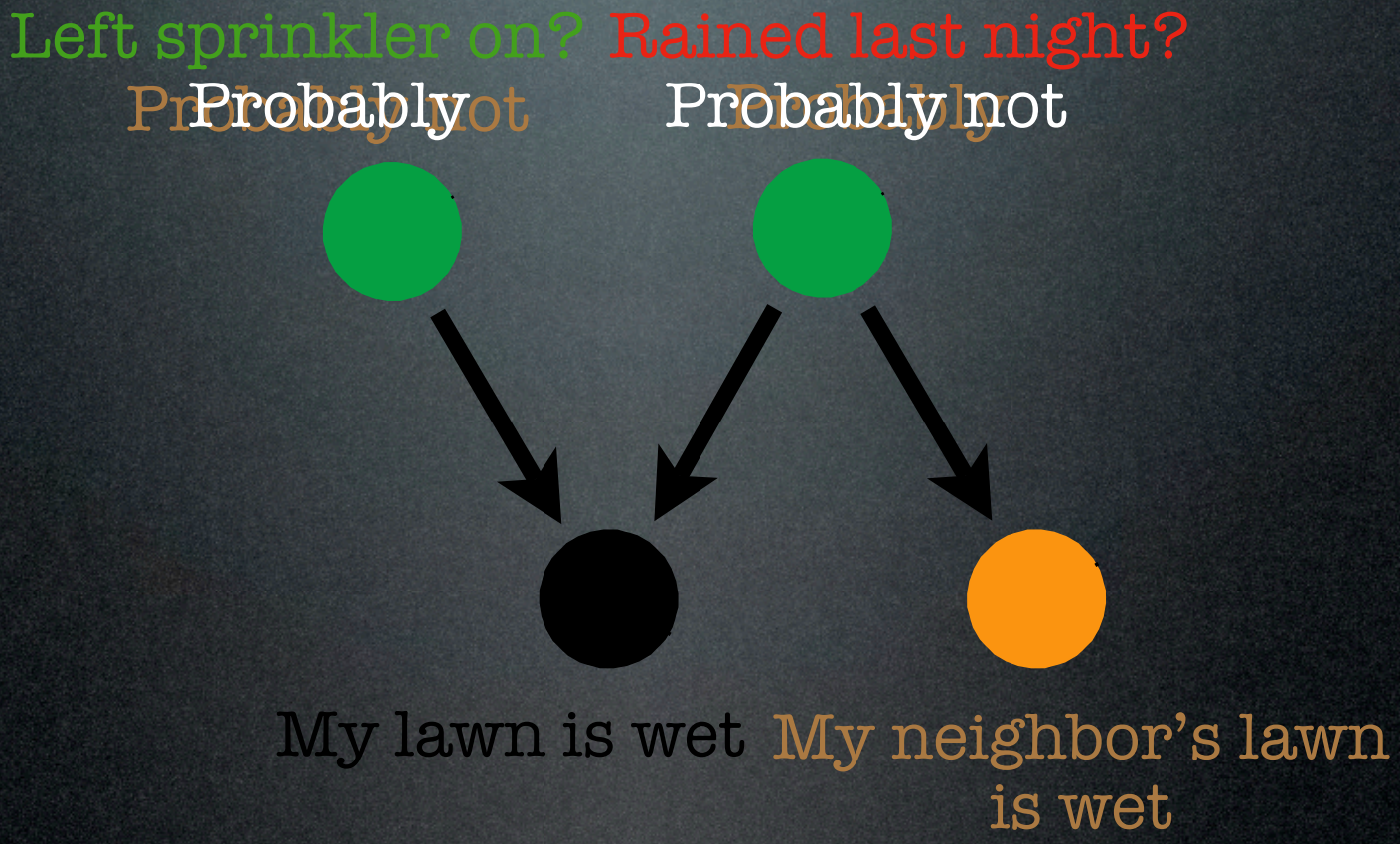
Basic Bayes



From: Schrater, P. and Sundareswara, R. (2006) "Theory and Dynamics of Perceptual Bistability", NIPS 2006

Explaining away

Did I leave the sprinkler on last night?



Pearl, Judea. (1988). Probabilistic reasoning in intelligent systems: networks of plausible inference (Rev. 2nd printing. ed.). San Mateo, Calif.: Morgan Kaufmann Publishers.

Outline

- Object perception: The problem of ambiguity
- Theory: Vision as statistical inference
- Behavior: Perceptual “puzzles”
- Mechanisms: Resolving ambiguity

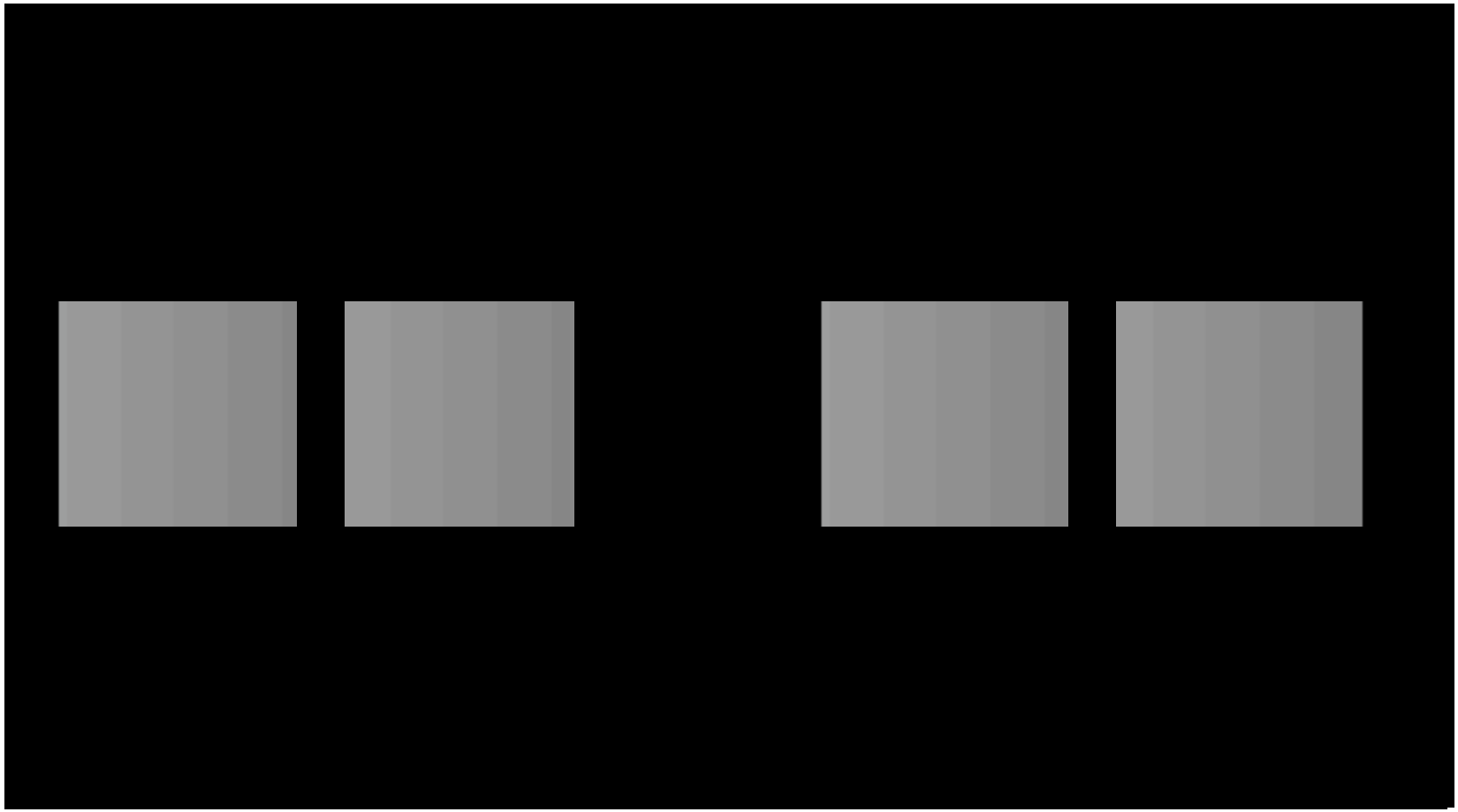
Resolving ambiguity without an “aha”

- Can be subjectively instantaneous
- or
- Gradual

Interactions between perception of geometry and material

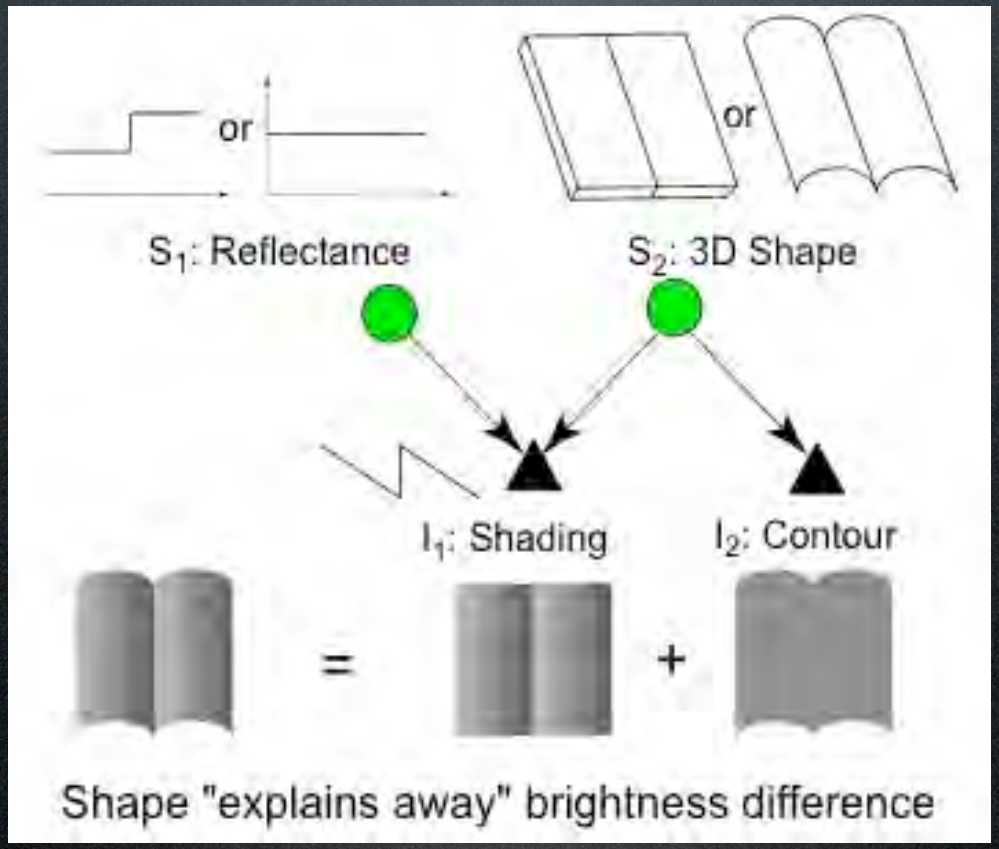
- No subjective ambiguity
- Subjective ambiguity

Lightness & shape



Knill & Kersten (1991)

Type of inference



- Apparent glossiness takes into account how highlights spread as a function of curvature
- Consistent with statistics of patterns of natural illumination

Matte or shiny?

Hartung & Kersten

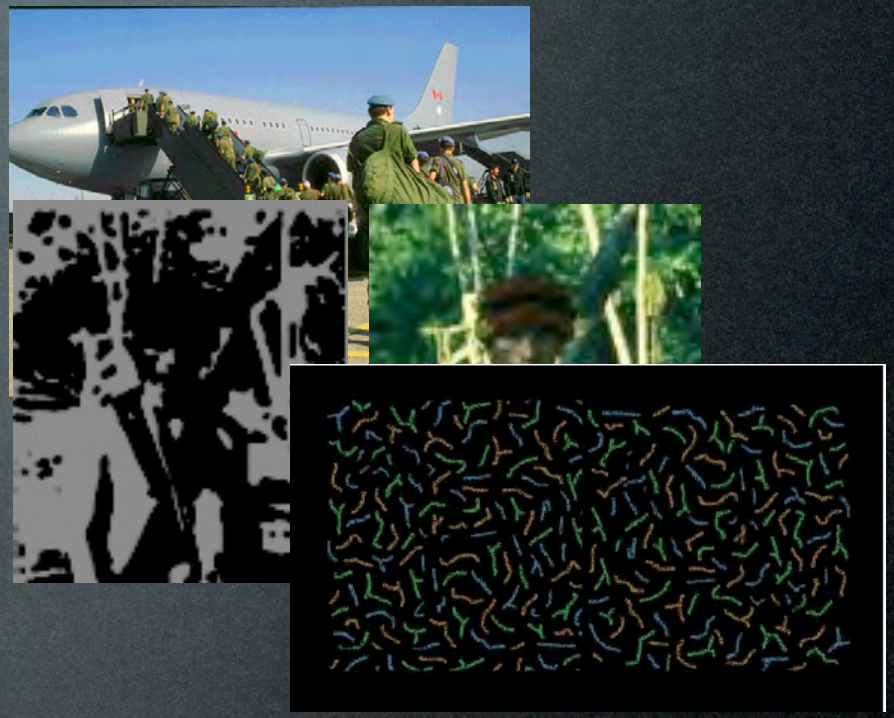
Learning a camouflaged object

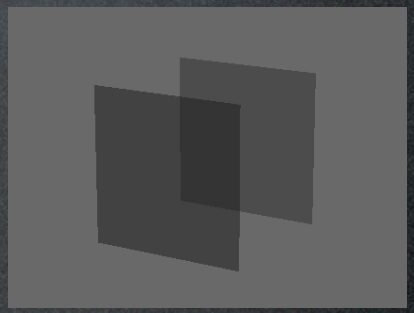
Training: each image is a different rendering

Test with novel camouflage

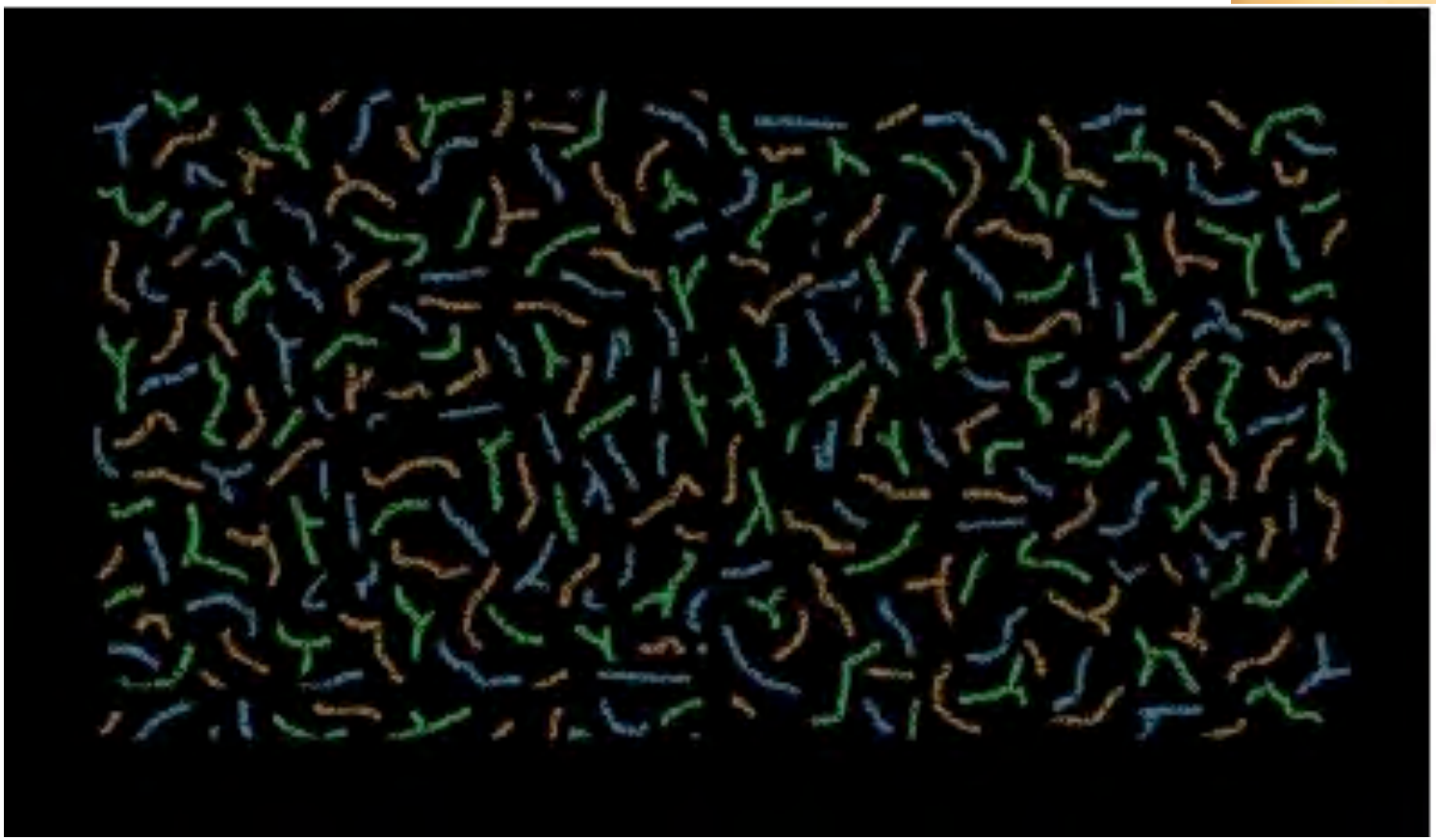
Some perceptual “aha’s”

- Change blindness
  http://www.cs.ubc.ca/~rensink/flicker/download/Airplane.mov
- Mooney
- Hidden figure
- Bistable perceptual organization



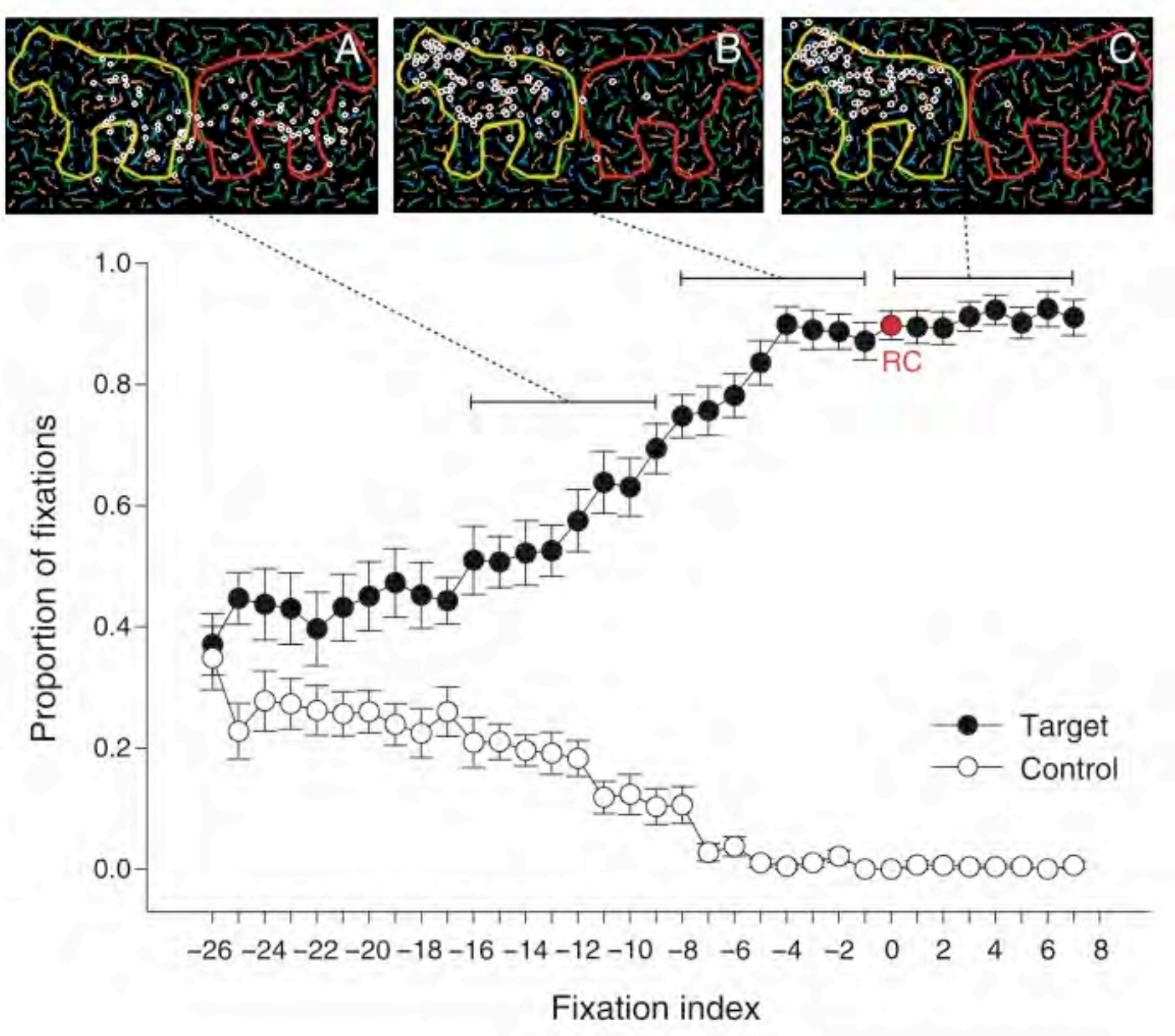


Hidden figure



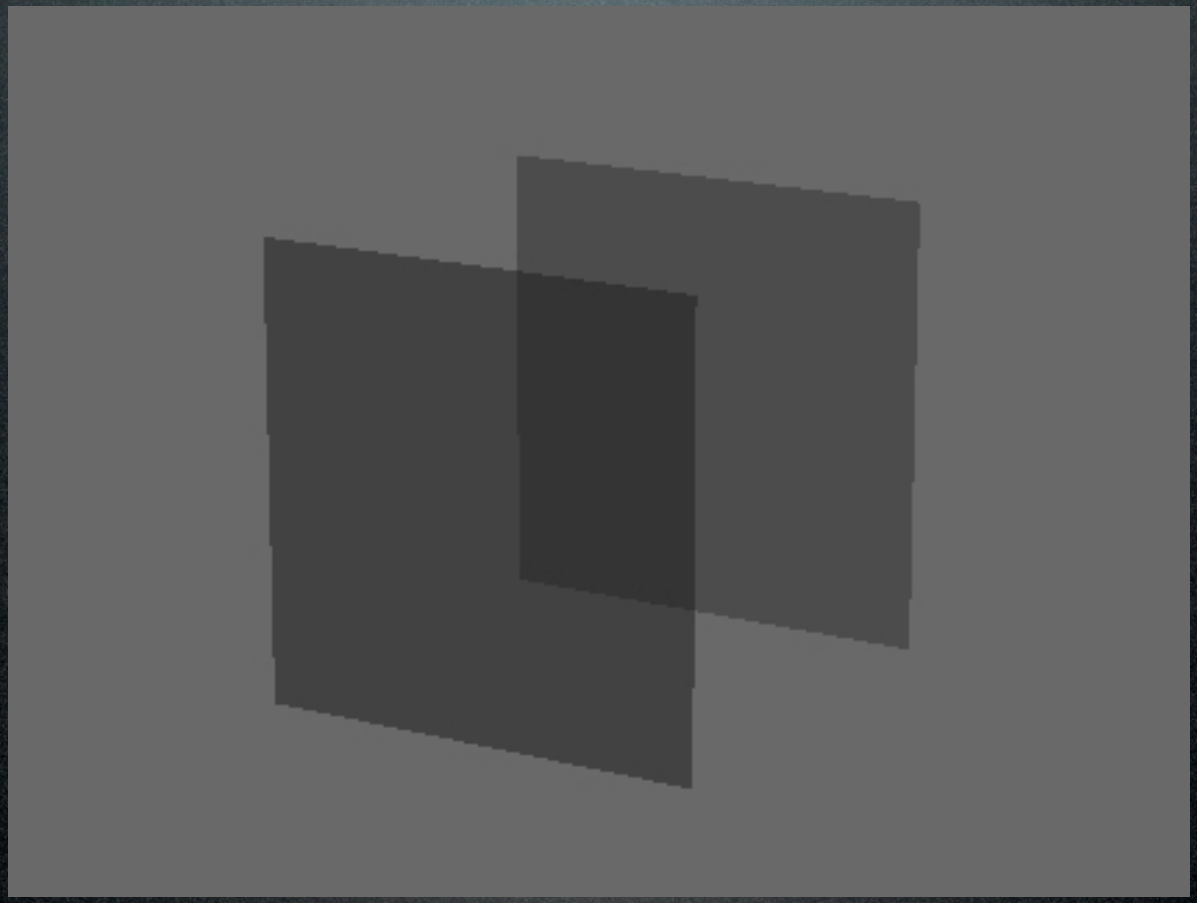
Linus Holm

Looking as if you know: Systematic object inspection precedes object recognition. Holm, Eriksson, & Andersson, Journal of Vision (2008) 8(4):14, 1–7

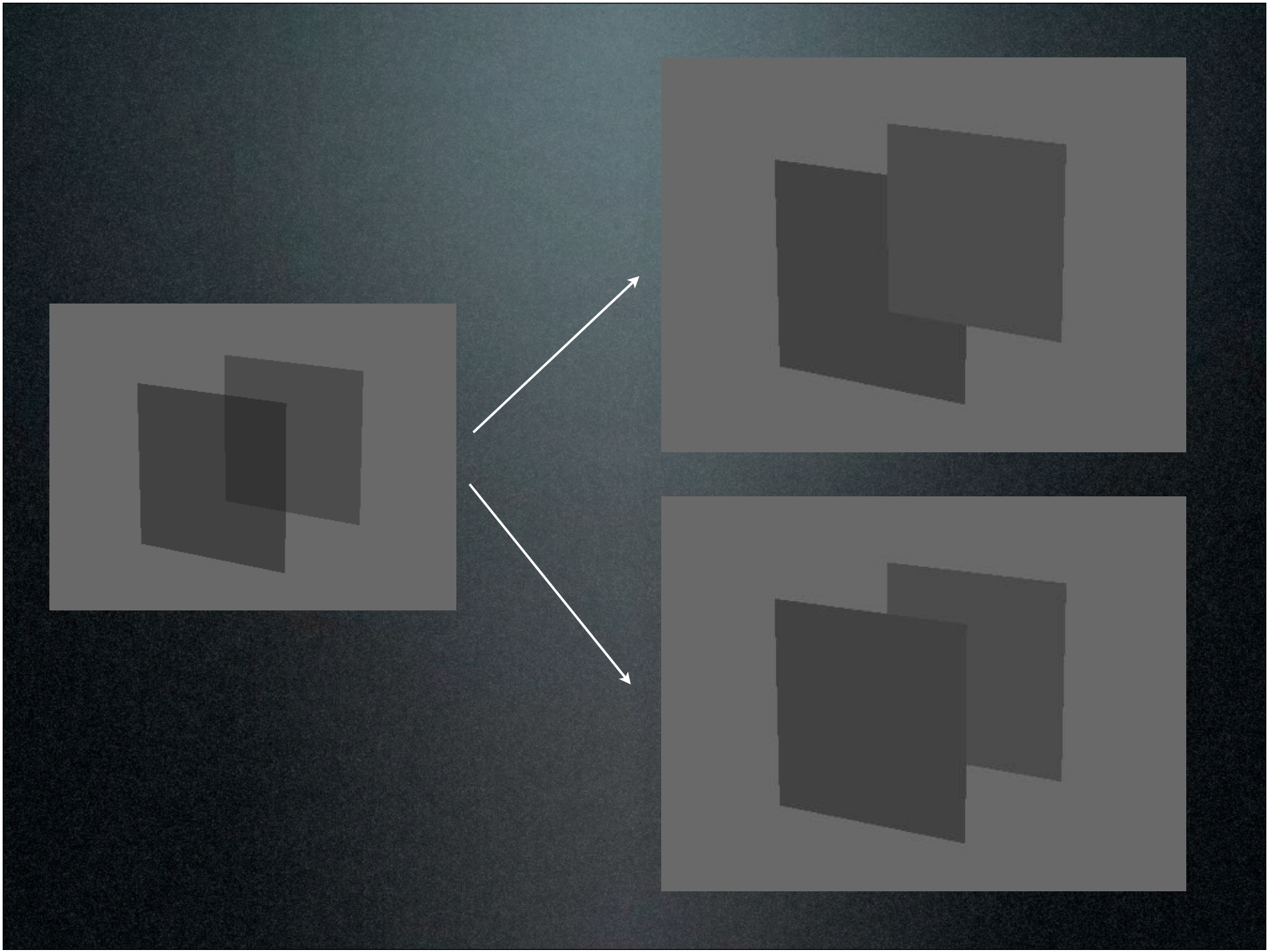


Looking as if you know: Systematic object inspection precedes object recognition. Holm, Eriksson, & Andersson, Journal of Vision (2008) 8(4):14, 1–7

Material and Geometrical Structure

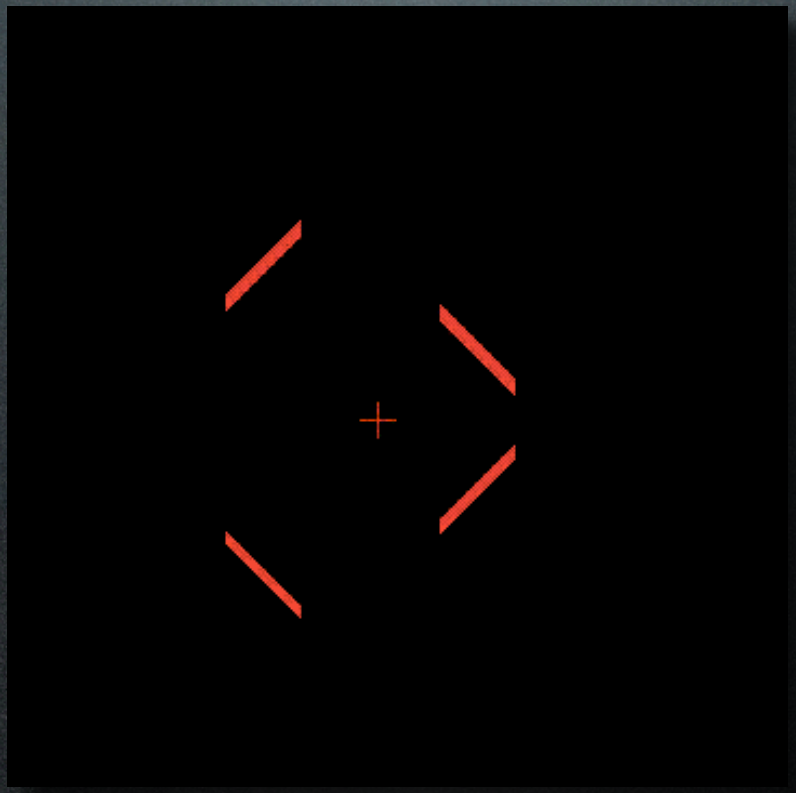


Kersten, D., Bülthoff, H. H., Schwartz, B., & Kurtz, K. (1992) Neural Computation, 4, 573-589.



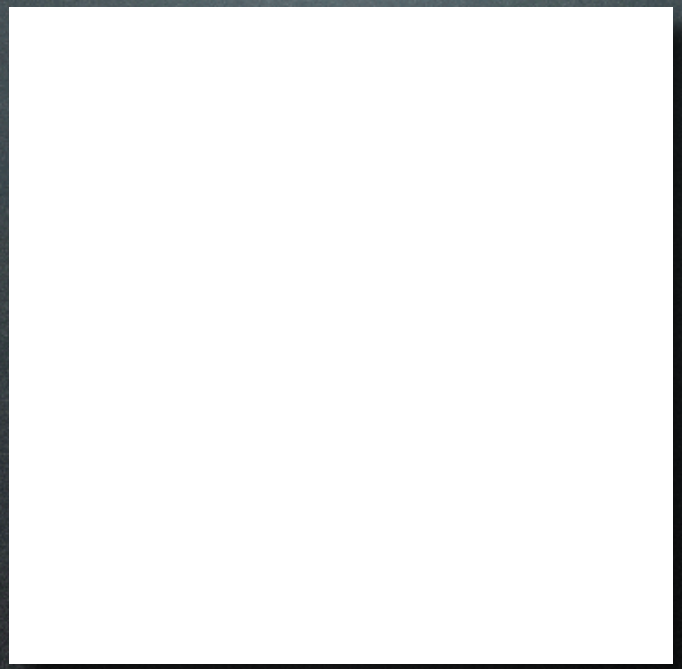
Perceptual organization of shape

Do you see a diamond moving horizontally?

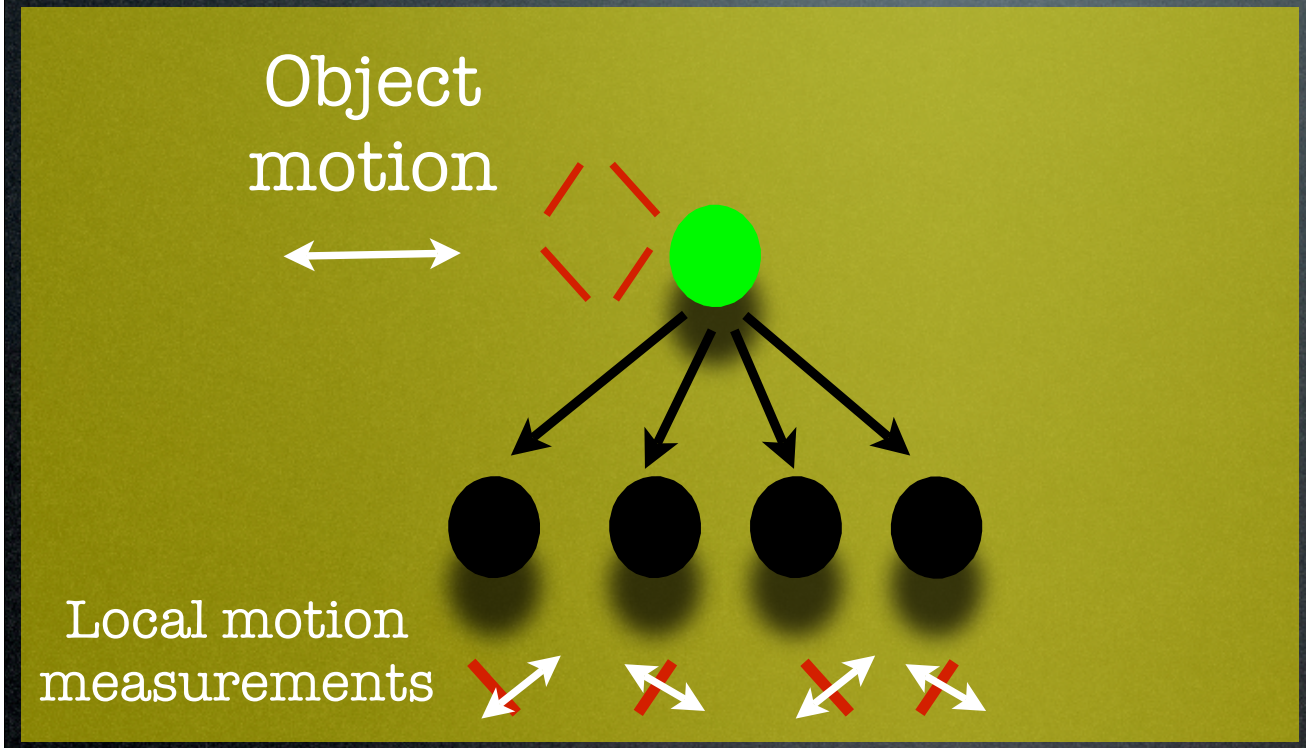


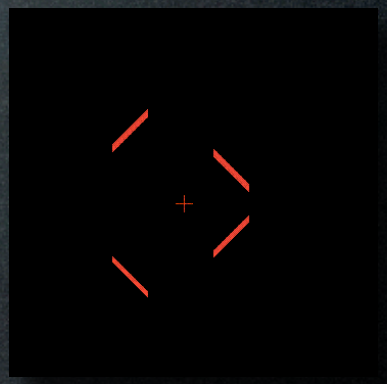
Lorenceau, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. Vision Res, 32(2), 263-273.

The “solution”



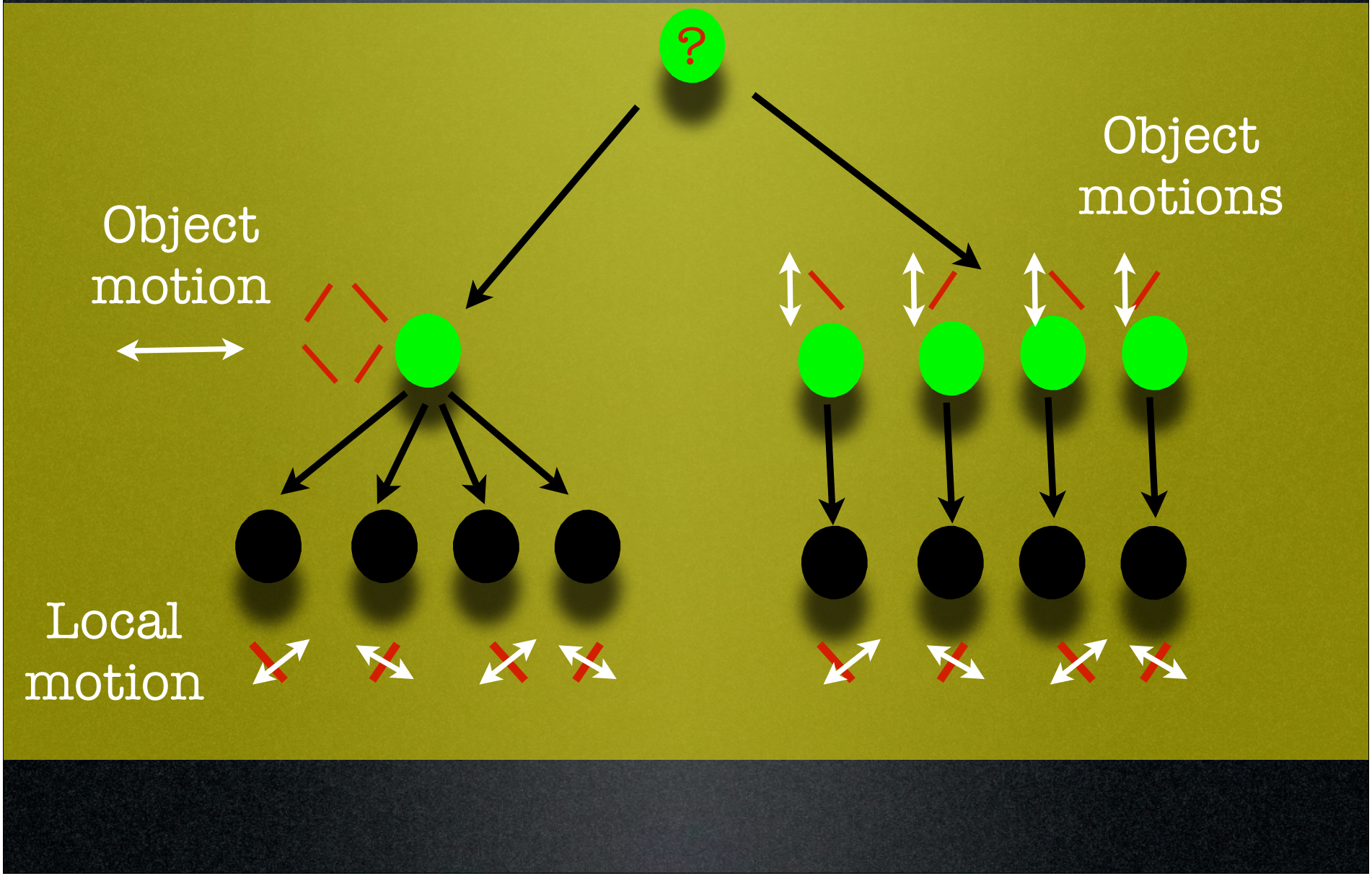
Computational problems: Local integration



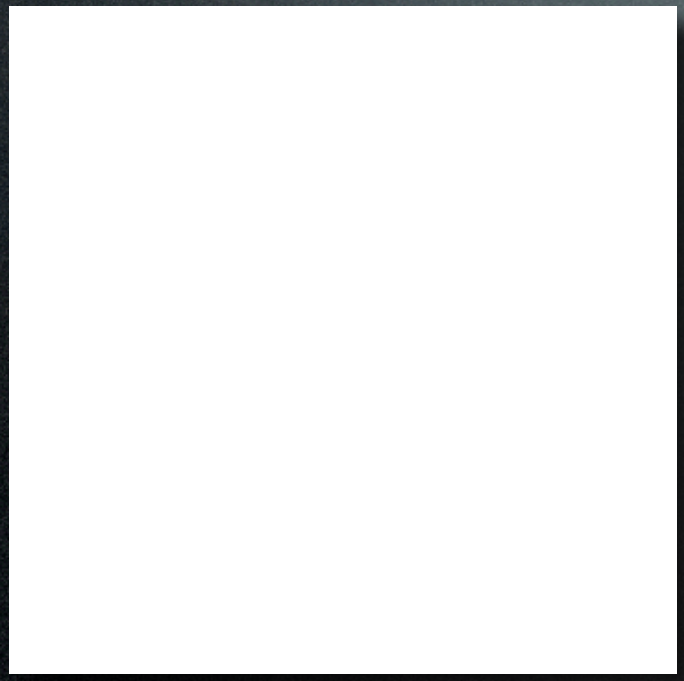


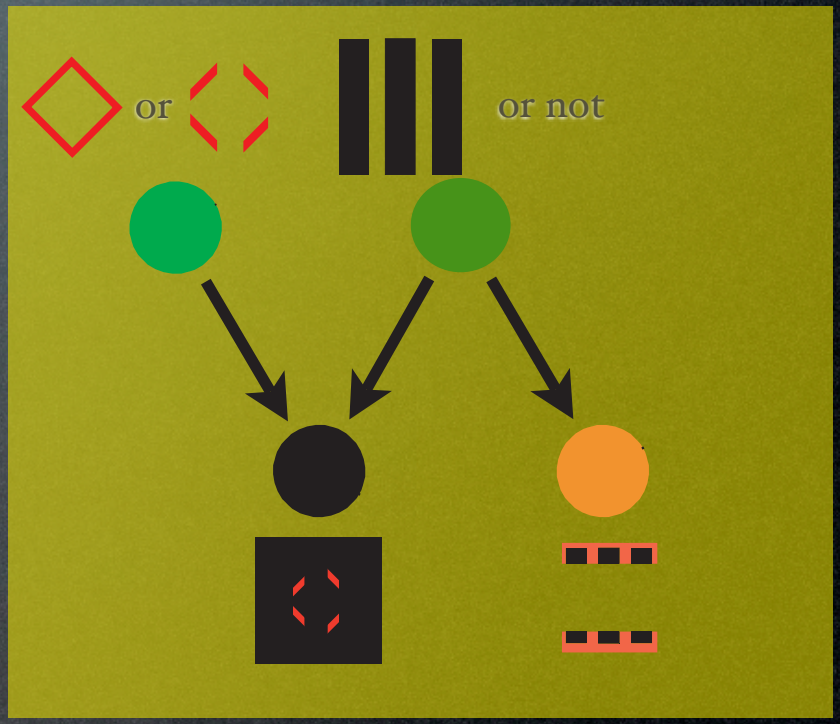
Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. Nat Neurosci, 5(6), 598-604.

Computational problems: Model selection



Computational problems: Explaining away

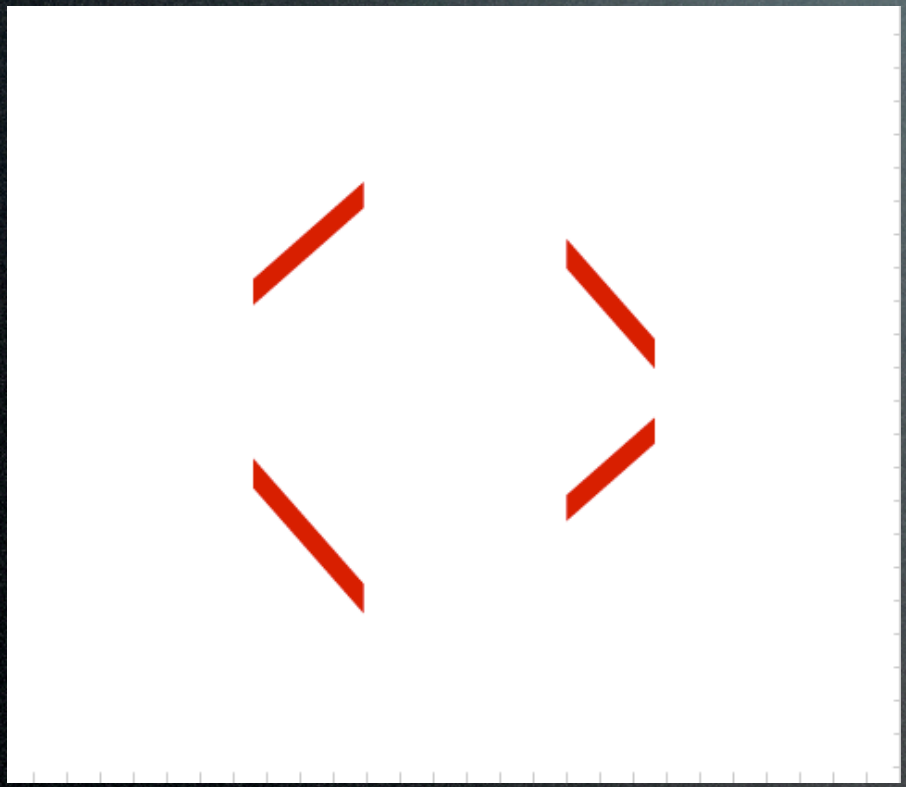


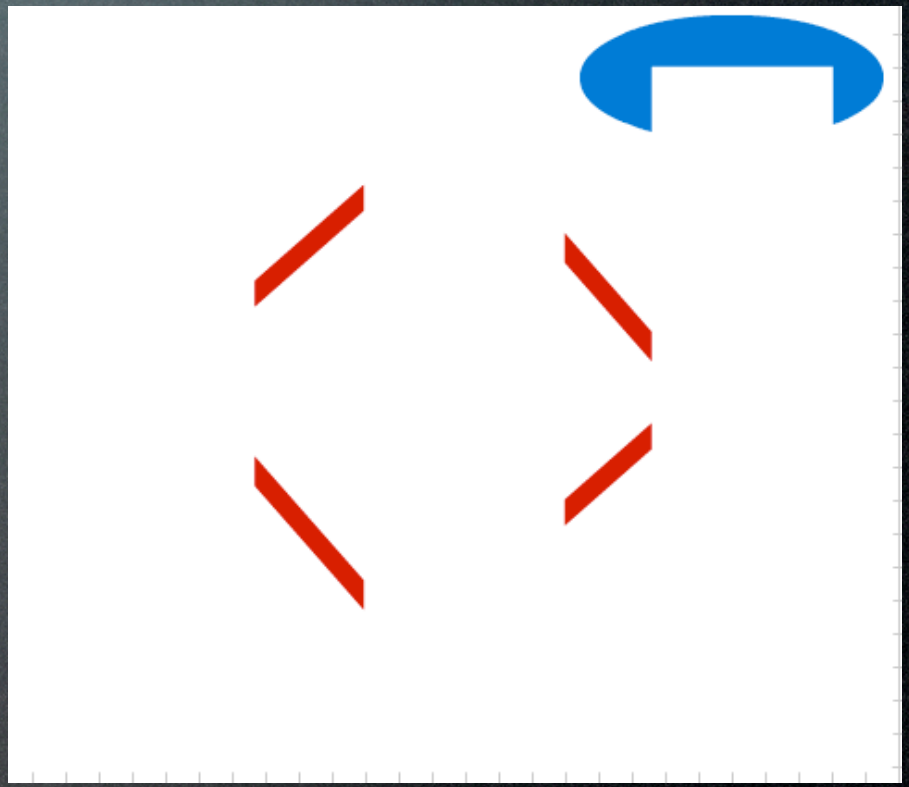


Auxiliary evidence for occlusion

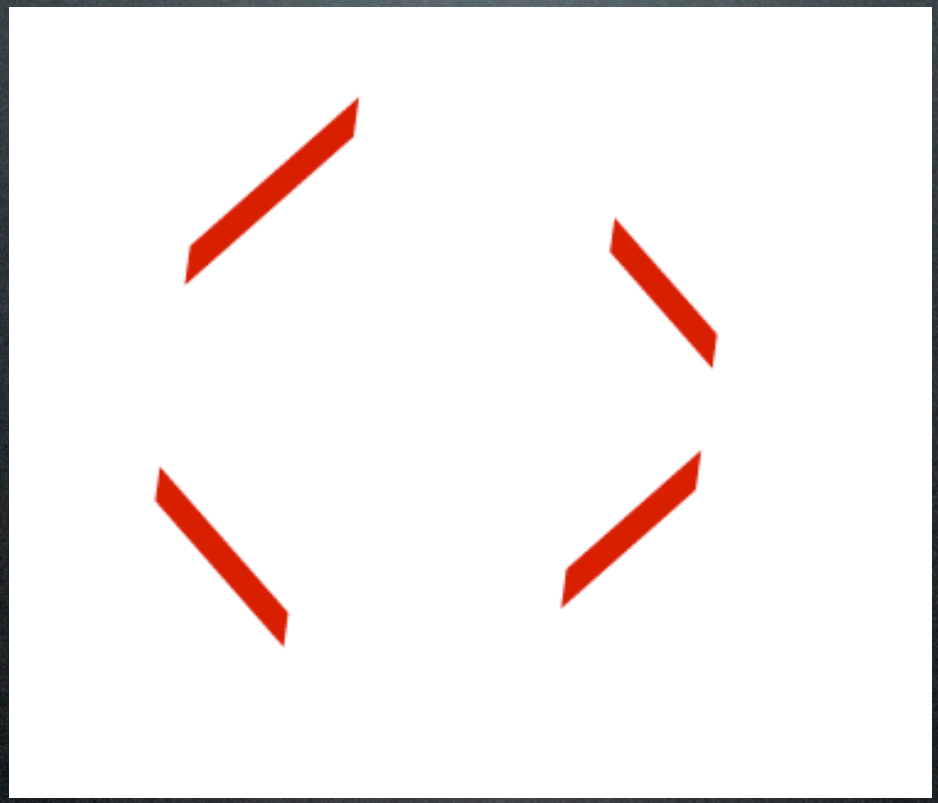
just for fun...another puzzle

A clue





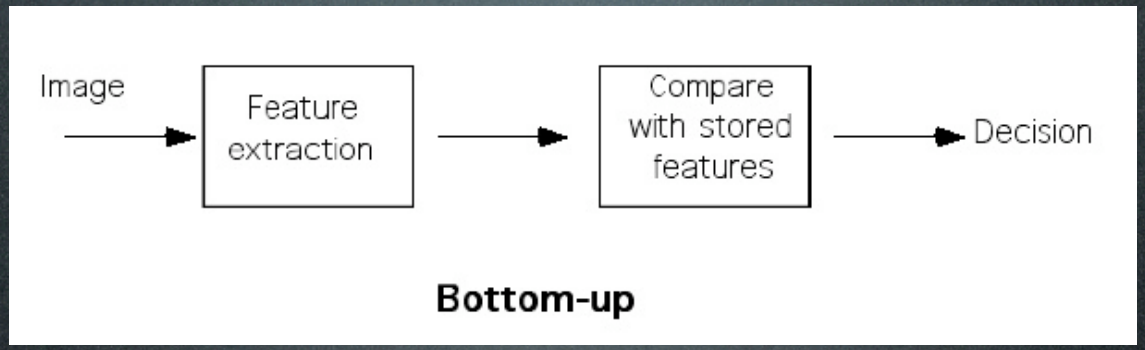
...and another

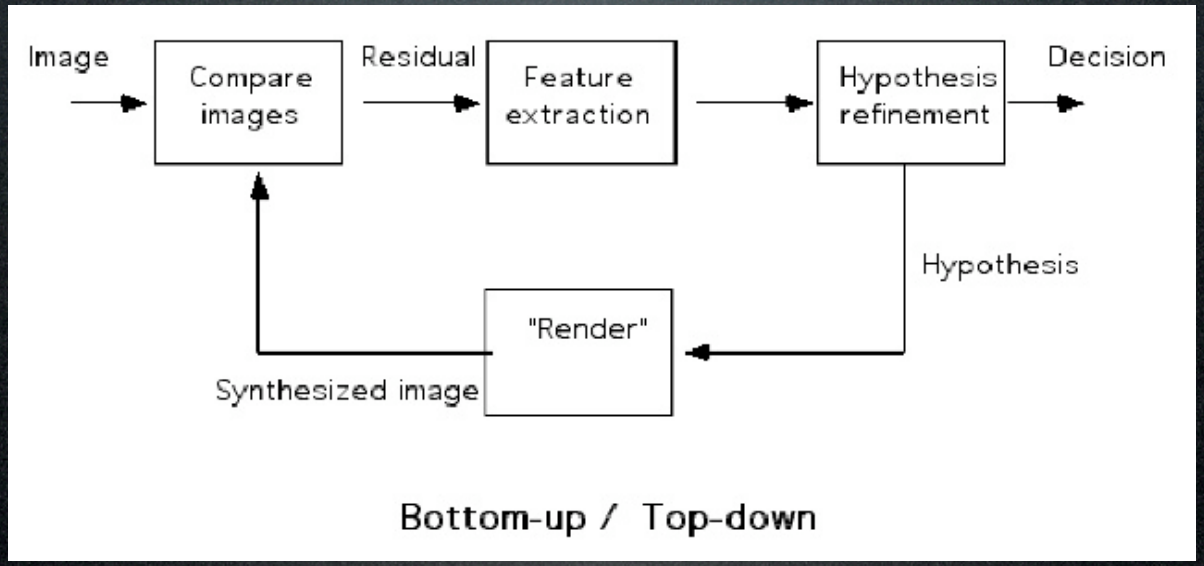


Outline

- Object perception: The problem of ambiguity
- Theory: Vision as statistical inference
- Behavior: Perceptual “puzzles”
- Mechanisms: Resolving ambiguity

"Analysis by synthesis"





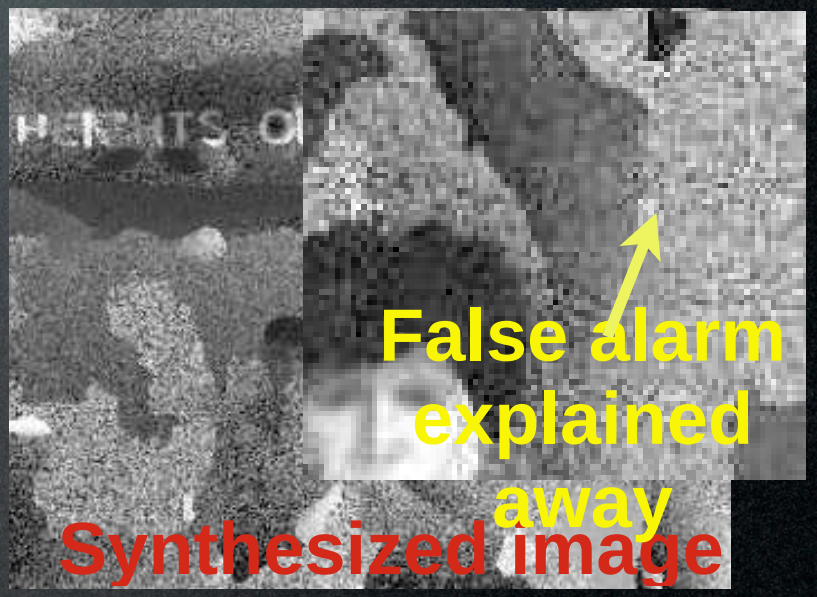
D. MacKay, U. Grenander, D. Mumford, Rao & Ballard ...

Analysis by synthesis particularly useful when:

- Ambiguity induced by clutter
- Transformations that are computationally difficult to do bottom-up, e.g.
  - orientation in 3D depth
  - articulations, e.g. scissors
  - occlusion
  - competing/interacting hypotheses

Three models: text, faces, texture

Input



Tu, Z., Chen, X., Yuille, A., & Zhu, S. (2005). Image Parsing: Unifying Segmentation, Detection and Recognition. IJCV, 63(2).

V1 activity & perceptual organization of shape

Use BOLD functional MRI to localize cortical activity that is correlated with the competing perceptual hypotheses of

- Coherent diamond vs. less coherent line fragments

Scott Murray, Dan Kersten, Bruno Olshausen, Paul Schrater and David Woods 2002. Proc Natl Acad Sci U S A, 99, 15164-15169.

V1: Feedforward model

Lots of theory...but most working models begin with simple assumptions:

- Banks of localized spatio-temporal filters (receptive fields)
  - Feedforward
    - with localized feedback for normalization/tuning

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. Nat Neurosci, 4(8), 819-825.

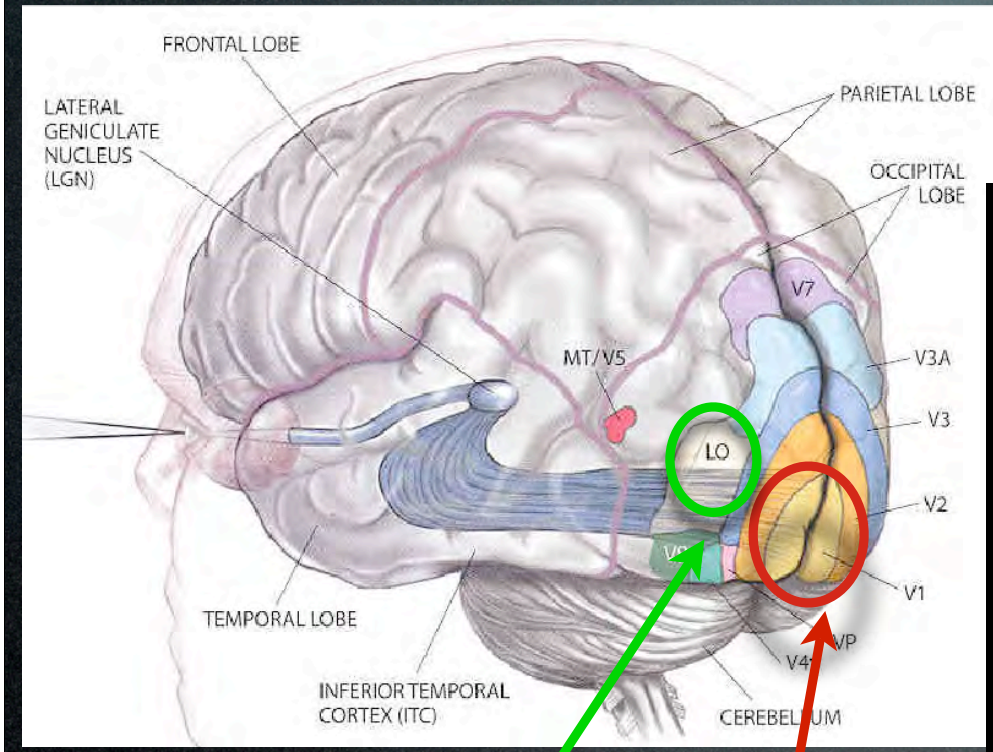
V1

- Feedforward processing by local, oriented filters would imply little or no effect of global structure, but we know...
  - within area connections
  - between area connections
  - longer range modulation of neural responses

Cf.

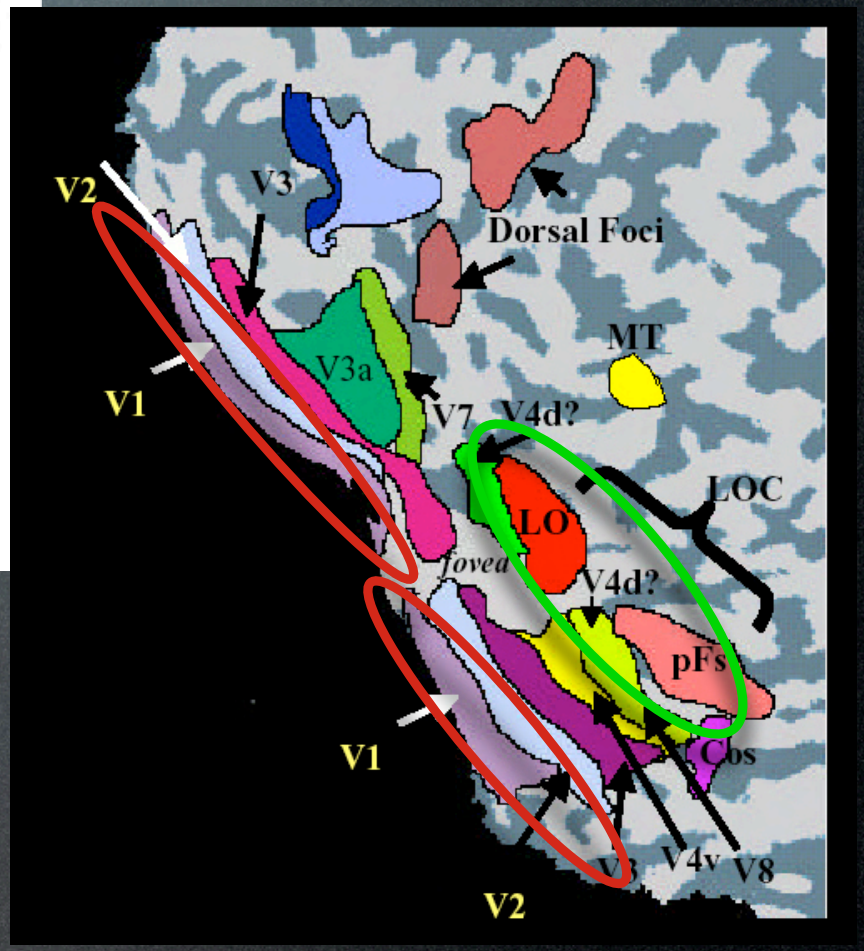
Bullier, J. (2001). Integrated model of visual processing. Brain Res Brain Res Rev, 36(2-3), 96-107.
Friston, K. (2005). A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci, 360(1456), 815-836.

Human V1



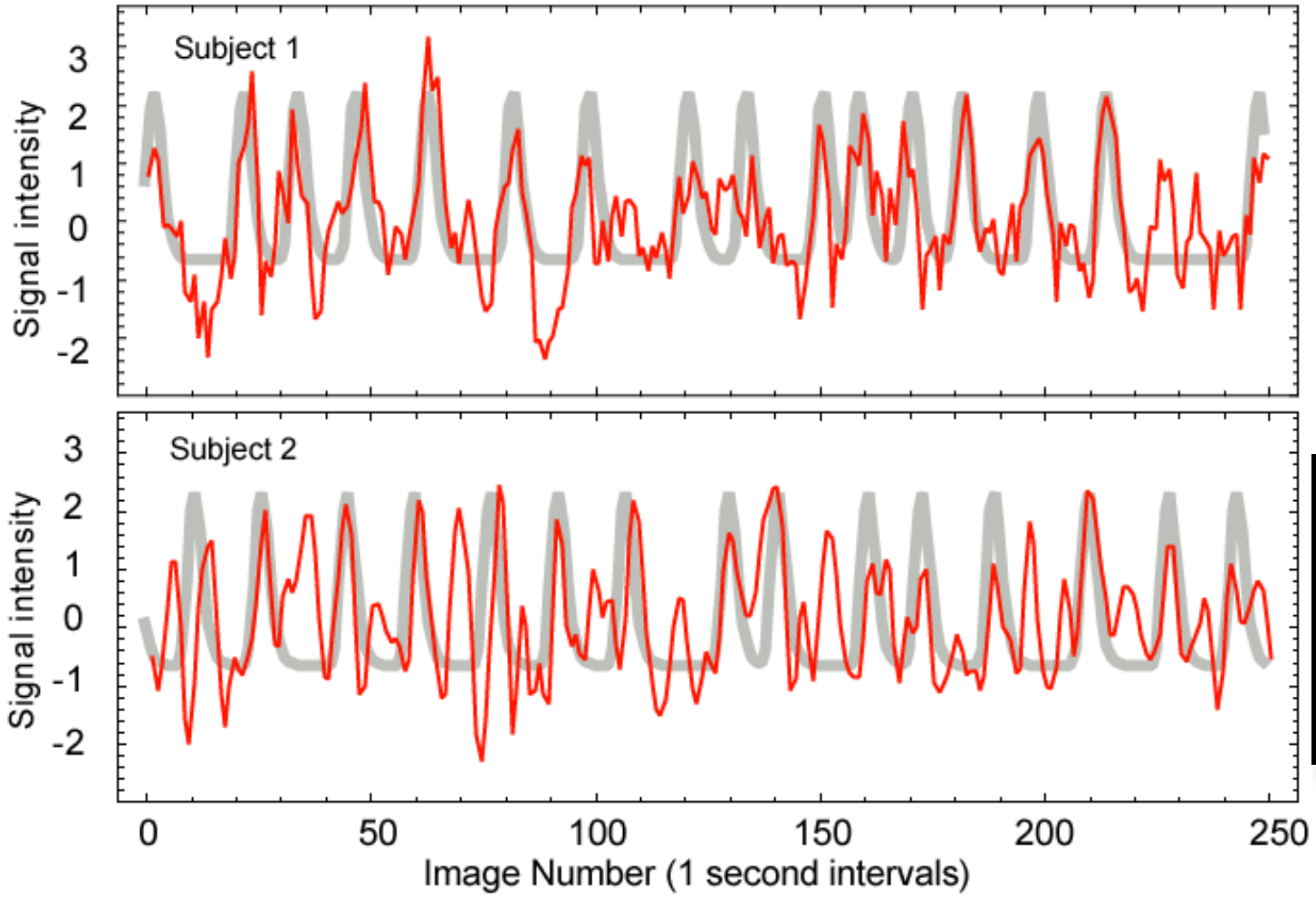
Whole shapes

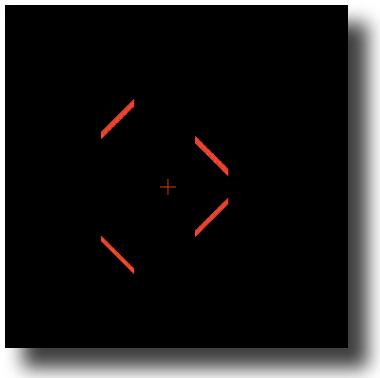
Local, oriented, moving edges



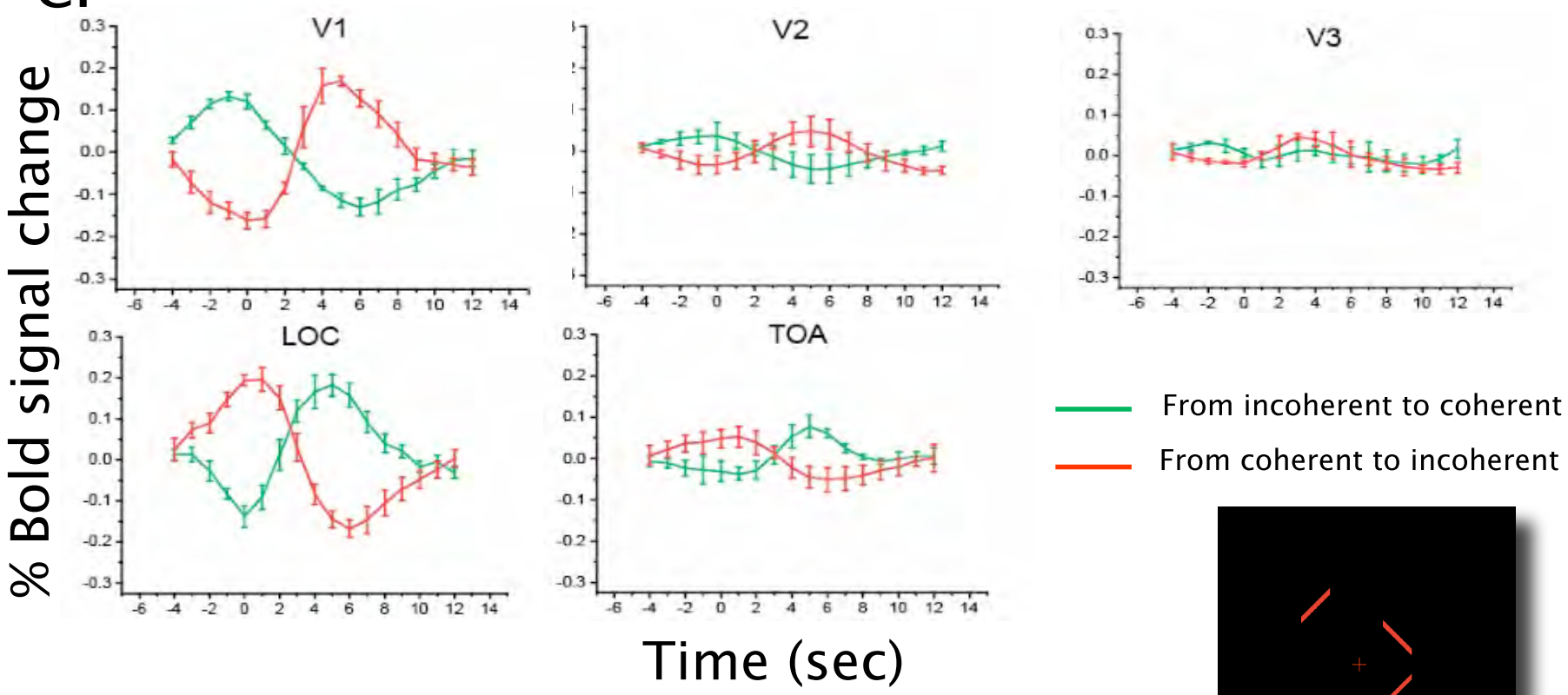
Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). Vision Res, 41(10-11), 1409-1422.

V1 response predicts behavior





C.

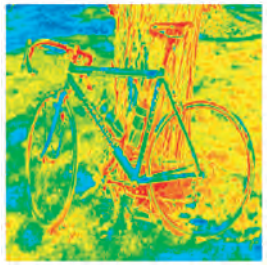


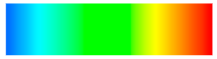
Fang F, Boyaci H, Kersten D, Murray SO (in press) Journal of Vision

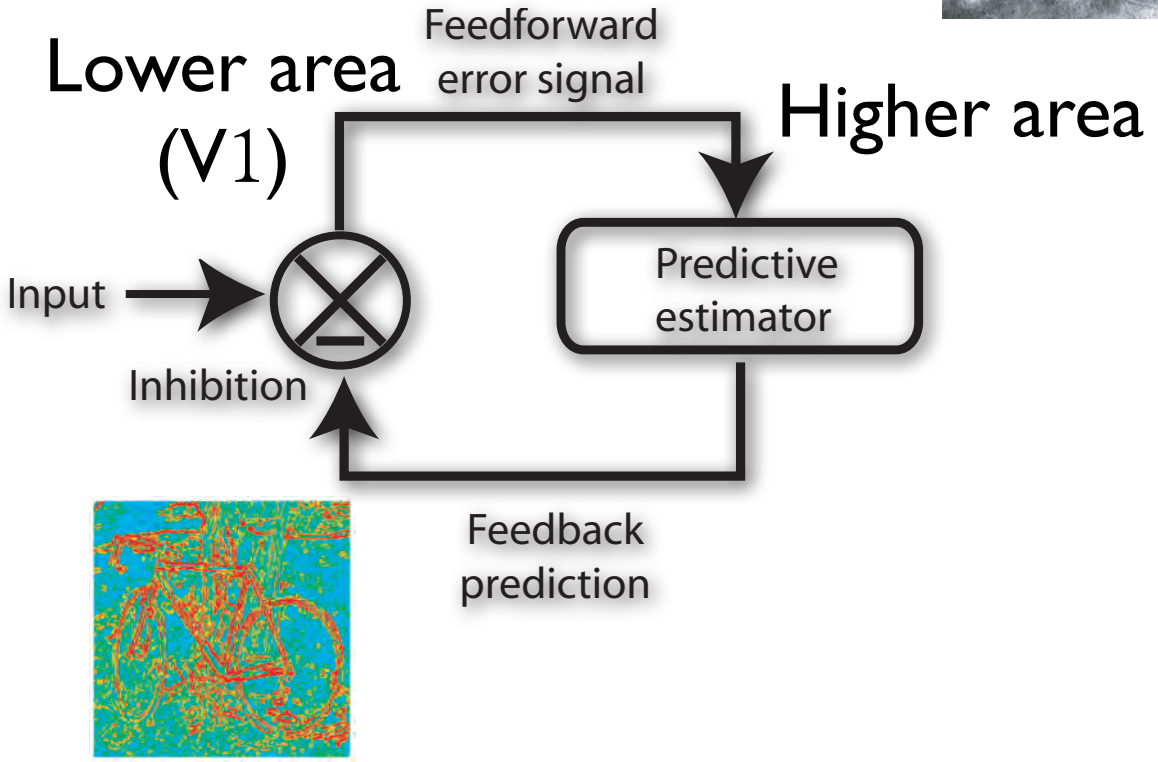
...story is not so simple

- Similar pattern of results:
  - Furl, van, Rijsbergen, Treves, Friston, & Dolan, 2007
  - Harrison, Stephan, Rees, & Friston, 2007
  - Summerfield et al., 2006
- ...but
  - Dumoulin & Hess, 2006
  - Lorenceau, Paradis, Lamirel, Poline, Artiges, Thirion & Caclin, VSS 2008
  - and some of our own (Jay Hegdé) results using Mooney and occluded images (Fang Fang)

Predictive coding

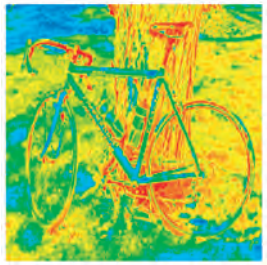


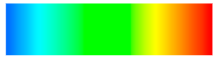




e.g. Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput, 9(4), 721-763.

Predictive coding





Lower area
(V1)

Feedforward
error signal

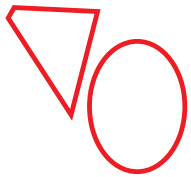
Higher area

Input

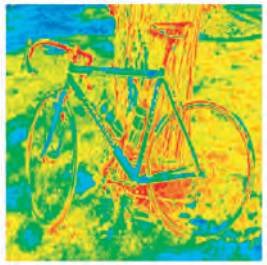
Predictive
estimator

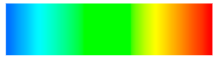
Inhibition

Feedback
prediction



Predictive coding





Lower area
(V1)

Feedforward
error signal

Higher area

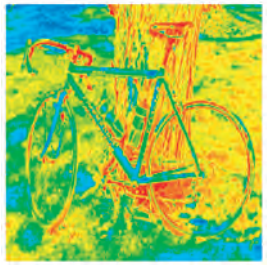
Input

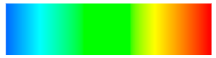
Predictive
estimator

Inhibition

Feedback
prediction

Predictive coding





Lower area
(V1)

Feedforward
error signal

Higher area

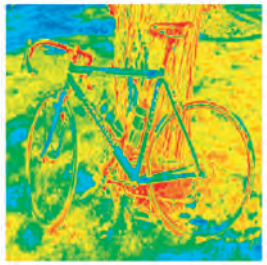
Input

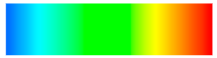
Predictive
estimator

Inhibition

Feedback
prediction

Predictive coding





Lower area
(V1)

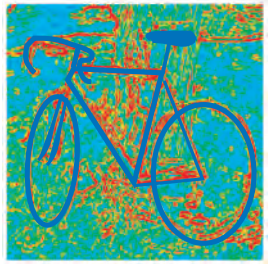
Feedforward
error signal

Higher area

Input

Predictive
estimator

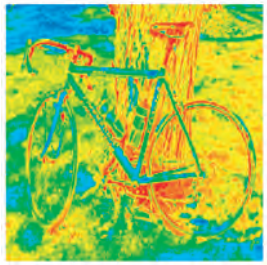
Inhibition

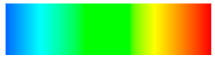


Feedback
prediction

Sparsification

"Stop gossiping"

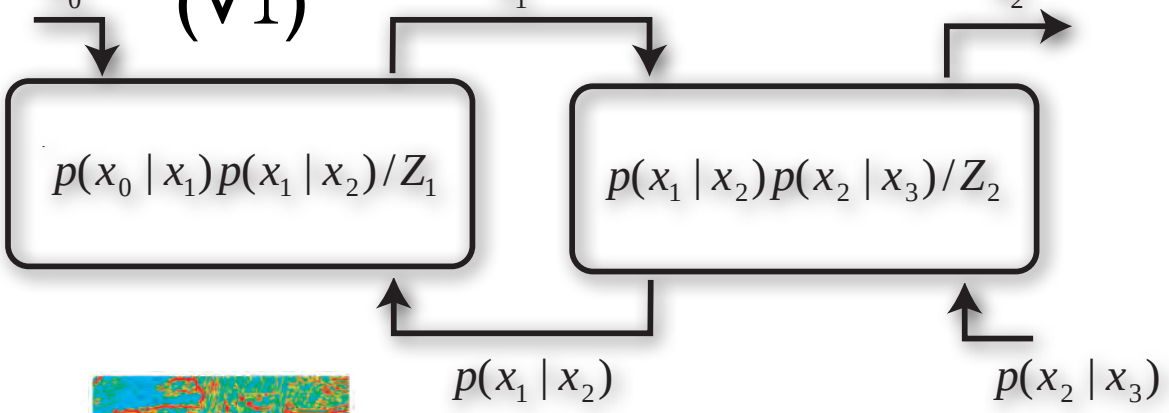


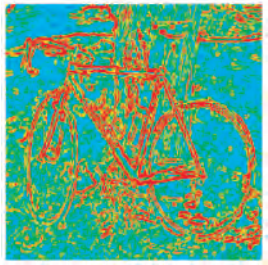


Lo Hi

Lower area (V1)

Higher area

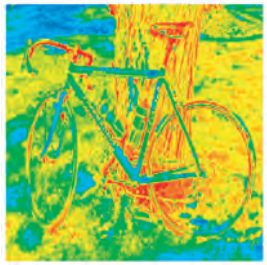


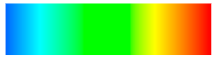


Lee & Mumford, 2003, JOSA

Sparsification

"Stop gossiping"





Lower area (V1)

Higher area

x_0 x_1 x_2

$p(x_0 \mid x_1)p(x_1 \mid x_2)/Z_1$

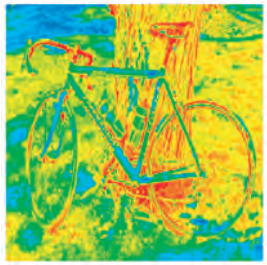
$p(x_1 \mid x_2)p(x_2 \mid x_3)/Z_2$

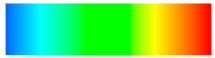
$p(x_1 \mid x_2)$

$p(x_2 \mid x_3)$

Sparsification

"Stop gossiping"





Lower area (V1)

Higher area

x_0

x_1

x_2

$p(x_0 \mid x_1)p(x_1 \mid x_2)/Z_1$

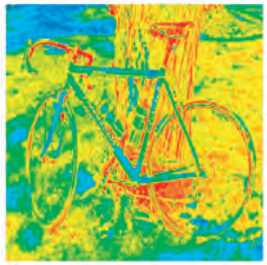
$p(x_1 \mid x_2)p(x_2 \mid x_3)/Z_2$

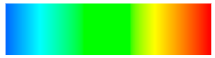
$p(x_1 \mid x_2)$

$p(x_2 \mid x_3)$

Sparsification

"Stop gossiping"





Lower area (V1)

Higher area

x_0

x_1

x_2

$p(x_0 \mid x_1)p(x_1 \mid x_2)/Z_1$

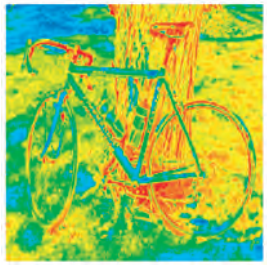
$p(x_1 \mid x_2)p(x_2 \mid x_3)/Z_2$

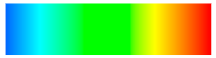
$p(x_1 \mid x_2)$

$p(x_2 \mid x_3)$

Sparsification

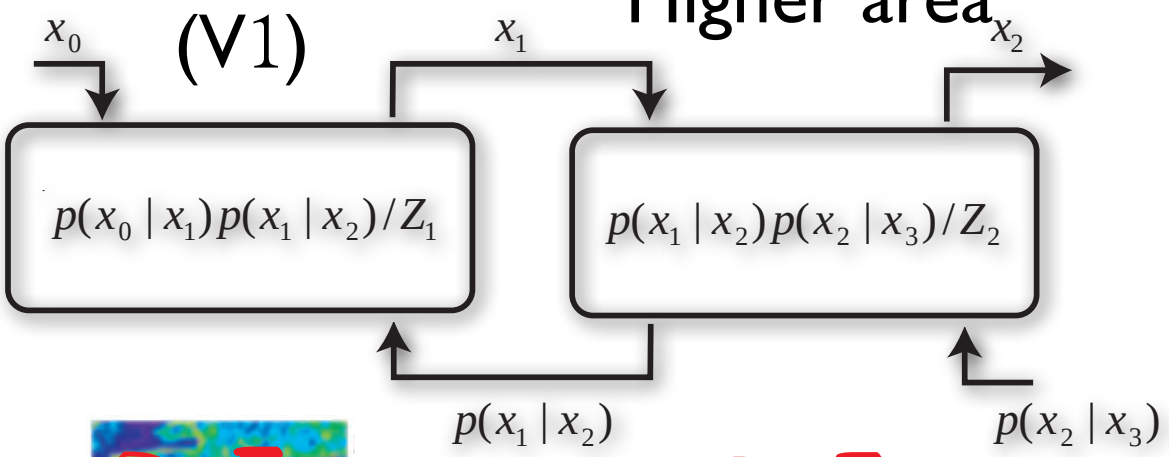
"Stop gossiping"

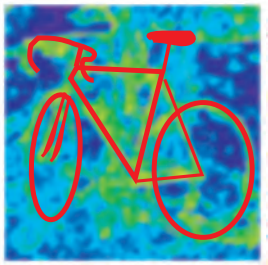




Lower area (V1)

Higher area





Summary

- Generative knowledge used to resolve ambiguity about objects
- Some instances of perceptual “aha’s” suggest generative mechanisms
- Requires an understanding of interactions between cortical areas

thanks...

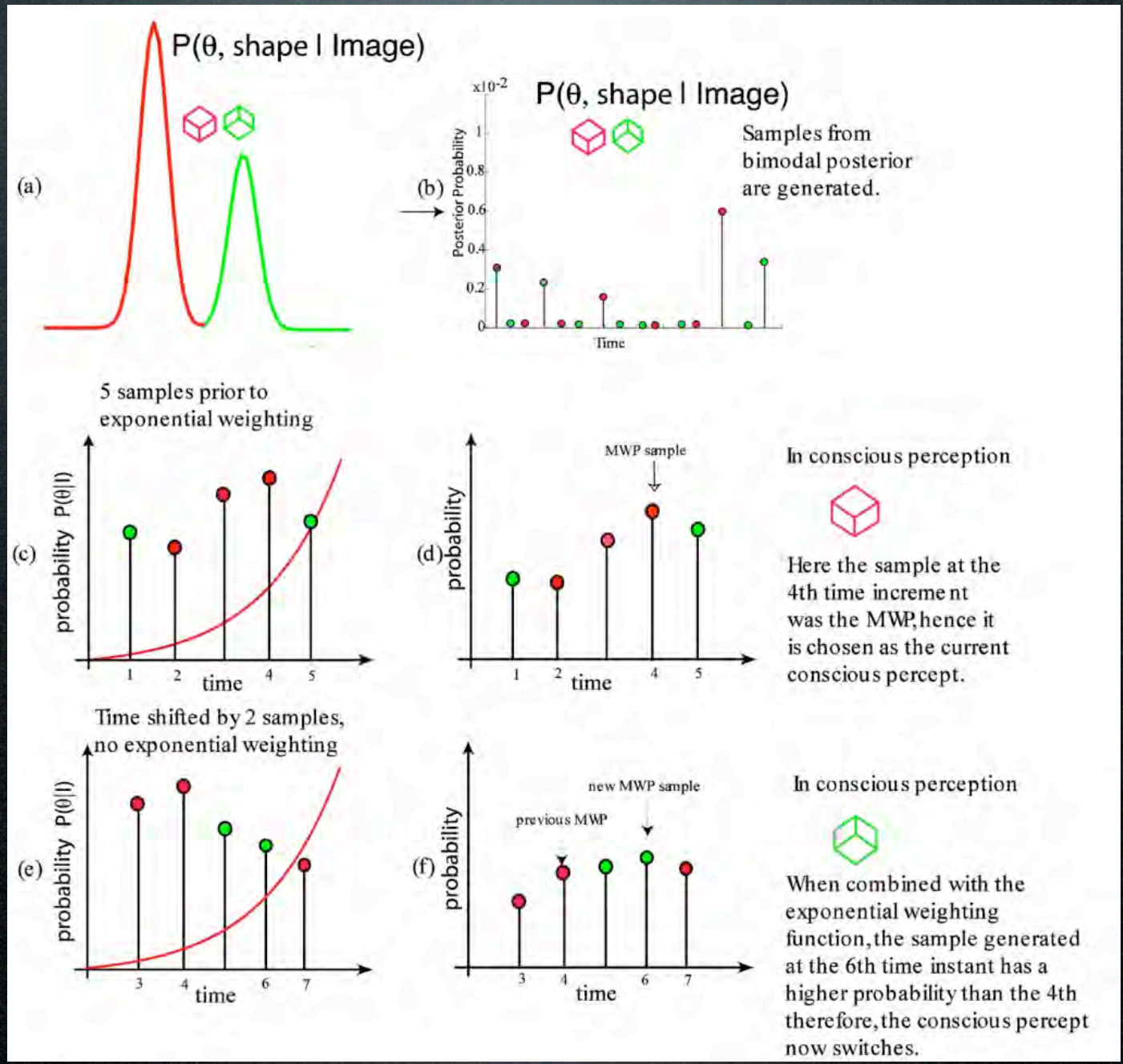
Current lab: Gina Albanese, Peter Battaglia, Charlie Benson, Huseyin Boyaci, Katja Doerschner, Jay Hegdé, Jennifer Schumacher, Bobby Shannon, Serena Thompson

Recent alumni: Scott Murray, Fang Fang

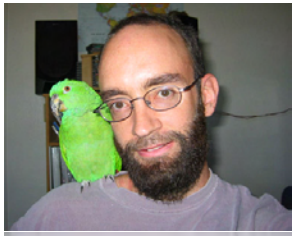
Collaborators: Sheng He, Paul Schrater, Cheryl Olman, Pascal Mamassian, Alan Yuille

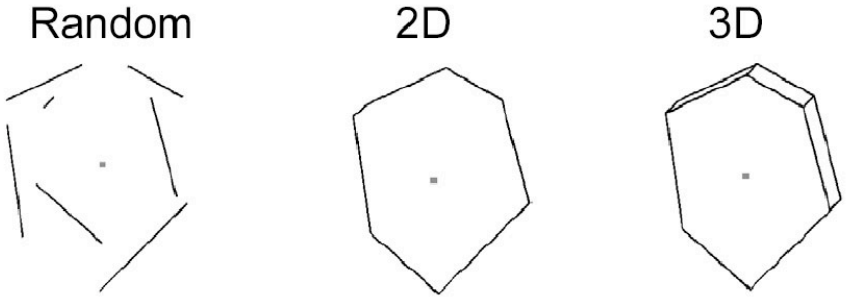
...and of course, NIH, NSF & ONR

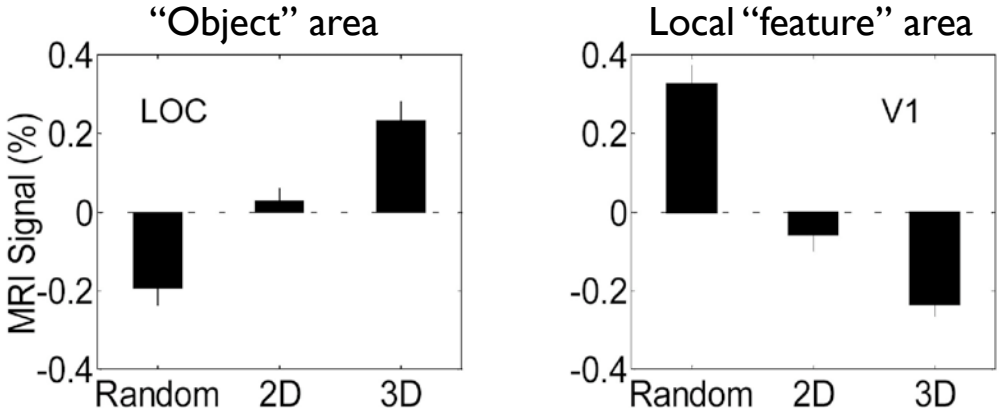
NIH R01 EY015261, NIH P41 RR08079, NSF SBR-9631682, N00014-05-1-0124


P(θ, shape | Image)
(a)
P(θ, shape | Image)
x10-2
Posterior Probability
(b)
Time
Samples from bimodal posterior are generated.
5 samples prior to exponential weighting
(c)
probability P(θ|I)
time
(d)
MWP sample
probability
time
In conscious perception
Here the sample at the 4th time increment was the MWP, hence it is chosen as the current conscious percept.
Time shifted by 2 samples, no exponential weighting
(e)
probability P(θ|I)
time
(f)
new MWP sample
previous MWP
probability
time
In conscious perception
When combined with the exponential weighting function, the sample generated at the 6th time instant has a higher probability than the 4th therefore, the conscious percept now switches.

Shape perception can reduce V1 activity







Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci U S A, 99, 15164-15169.